



A permutation test for the race model inequality

Gondan, Matthias

Published in:
Behavior Research Methods

DOI:
[10.3758/BRM.42.1.23](https://doi.org/10.3758/BRM.42.1.23)

Publication date:
2010

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Gondan, M. (2010). A permutation test for the race model inequality. *Behavior Research Methods*, 42(1), 23-28.
<https://doi.org/10.3758/BRM.42.1.23>

A permutation test for the race model inequality

MATTHIAS GONDAN

University of Regensburg, Regensburg, Germany

When participants are asked to respond in the same way to several stimulus identities, responses are often observed to be faster if two stimuli are presented simultaneously as opposed to when a single stimulus is presented (redundant signals effect; Miller, 1982). An important issue of such experiments is whether the observed redundancy gains can be explained by parallel processing of the two stimuli in a race-like fashion. To test the parallel processing model, Miller derived the well-known race model inequality which has become a routine test for behavioral data in experiments with redundant signals. Several statistical procedures have been used for testing the race model inequality. However, the commonly employed procedure does not control the Type I error. In this article a permutation test is described that keeps the Type I error at the desired level. Simulations show that the power of the test is reasonable even for small samples. The scripts discussed in this article may be downloaded as supplemental materials from <http://brm.psychonomic-journals.org/content/supplemental>.

Race Model Inequality

The redundant signals paradigm is a standard setup in experimental psychology. In this setup, participants are asked to respond in the same way to stimuli from two sources, for example, auditory and visual signals (A, V). It is frequently observed that responses are faster if two bimodal target stimuli are presented simultaneously (AV) rather than single, unimodal stimulus presentation. This is called the *redundant signals effect* (e.g., Miller, 1982).

There are different explanations that account for the redundant signals effect, the most important being the race model and the coactivation model. According to the race model (Miller, 1982) or parallel, first-terminating model (Townsend & Ashby, 1983), redundancy gains are explained by means of “statistical facilitation” (Raab, 1962). The time to detect and respond to a single target stimulus varies between trials and follows a statistical distribution. Whenever two target stimuli are presented, the stimulus faster processed in a given trial determines the response (i.e., this stimulus wins the race). As the processing time distributions for the two stimuli usually overlap, slow processing times are eliminated. Consequently, the mean response times (RTs) for redundant stimuli are shorter than those for the single stimuli. Instead of separate processing, the second class of explanation proposes an integrated processing of the two signals. For example, Miller (1982) proposed that the activation induced by both stimuli adds up, and that the summed activation enables faster response initiation (“coactivation model”).

In order to distinguish between statistical facilitation (race model) and integrated processing (coactivation model), Miller (1982) derived the well-known race model inequality, stating that the cumulative RT distribution for

the redundant stimuli (F_{AV}) never exceeds the sum of the RT distributions for the unimodal stimuli (F_A, F_V).

$$F_{AV}(t) \leq F_A(t) + F_V(t) \quad (1)$$

If the race model inequality is violated in a given data set, parallel processing cannot account for the redundancy gain. Most researchers would then accept the alternative type of explanation (i.e., integrated processing of the two target stimuli) for the redundancy gain.

For testing Inequality 1 in single participants, nonparametric tests have been proposed by Miller (1986), G. Maris and Maris (2003), and recently by Vorberg (2008). In most experiments, however, data is collected from a group of participants. In the next section, I briefly describe the statistical procedure typically used for testing Inequality 1 in group data, and I summarize problems with Type I error probability related to this standard procedure. Then I describe a permutation test which keeps the Type I error rate at the nominal significance level while having reasonable power to detect violations of the race model.

Paired *t* test

The race model can be tested by applying paired *t* tests to the inverse cumulative distribution functions (CDFs) of RTs for A, V, and AV at prespecified percentiles (Miller, 1982). Let f and g be strictly monotonically increasing functions, with $f(x) \leq g(x)$, for all x . Then, for the inverse functions, the reverse relation holds, $f^{-1}(y) \geq g^{-1}(y)$, for all y . Inequality 1 is thus restated, using the inverse cumulative distributions (percentile functions),

$$F_{AV}^{-1}(p) \geq F_{A+V}^{-1}(p), \quad (2)$$

with F_{A+V} denoting the sum of the two distributions $F_A(t) + F_V(t)$. A test of Inequality 2 is performed as fol-

lows (Ulrich, Miller, & Schröter, 2007): First, estimates \hat{G}_A , \hat{G}_V , and \hat{G}_{AV} of the distribution functions F_A , F_V , and F_{AV} are determined. Second, $\hat{G}_{A+V}(t) = \hat{G}_A(t) + \hat{G}_V(t)$ is computed for each participant. Third, percentile values $\hat{G}_{AV}^{-1}(p)$ and $\hat{G}_{A+V}^{-1}(p)$ are estimated at prespecified probabilities p . Finally, for each pair of percentiles, a paired t test is computed to compare $\hat{G}_{AV}^{-1}(p)$ to $\hat{G}_{A+V}^{-1}(p)$. If $\hat{G}_{AV}^{-1}(p)$ is significantly smaller than $\hat{G}_{A+V}^{-1}(p)$ at any p , the race model is rejected. Using simulations with log-normal RT distributions, Miller (1982, note 3) showed that comparison of the percentile differences $\hat{G}_{AV}^{-1}(p) - \hat{G}_{A+V}^{-1}(p)$ with a paired t test yielded robust Type I error rates under the race model. Ulrich et al. provided a computer program, Pascal and MATLAB code, for this computation.

Problems arise if the number of trials per condition, n , is too small to get unbiased estimates of the percentile functions, and if Inequality 2 is tested at multiple percentiles (Kiesel, Miller, & Ulrich, 2007). When testing the race model inequality as described above, three different distributions for $F_{AV}(t)$, $F_A(t)$, and $F_V(t)$ are estimated, each of which might be biased in small samples (Gilchrist, 2000). Kiesel et al. ran several Monte Carlo simulations to explore how biases in percentile estimation affect the race model inequality when computing paired t tests. Their results revealed that, for small samples, estimating the percentile functions in the single and redundant target conditions results in systematic biases against the race model. Kiesel et al. recommended at least $n = 20$ trials per condition; for cases with fewer than 20 trials, a computer program (RMIBias) can be used to estimate, based on simulation, the bias for a given percentile.

If the race model holds, Inequality 1 holds for all t . At which values of t should the inequality actually be tested? For the paired t test, Kiesel et al. (2007) showed that Type I error accumulation can be fairly substantial if t tests are carried out at multiple percentiles. To minimize the Type I error inflation, Kiesel et al. proposed that researchers restrict the percentile range for the test of the race model, for example, to percentiles 10 to 35. Second, the Type I error probability for each single t test can be adjusted to a level that keeps the overall Type I error rate at the desired 5% level; and third, researchers might postulate that t tests become simultaneously significant at multiple percentile points in order to reject the race model. Kiesel et al. provided a computer program (RMIErrors) that, again on the basis of simulations, estimates the overall Type I error probability, depending on how many t tests are done at which percentile values and at which nominal Type I error level.

Permutation Test for Multiple Percentiles

Permutation tests provide elegant ways to control for the overall Type I error in correlated significance tests. Consider a race model with maximally possible redundancy gain, such that in the lower percentile range, Inequality 1 becomes an equality, $F_{AV}(t) = F_A(t) + F_V(t)$ (e.g., G. Maris & Maris, 2003). Suppose that data have been collected for $N = 20$ participants, and Inequality 1 is tested at one single t . In a standard t test for paired samples, the test statistic would be $t_{\text{obs}} = (\sum_i d_i / N) / (\sigma_d / \sqrt{N})$ with N denoting the sample size, and $d_i = \hat{G}_{AV}(t) - \hat{G}_A(t) - \hat{G}_V(t)$ denoting the observed race model “difference” for participant i . Positive d_i reflect a

violation of the race model prediction. If d_i follow a normal distribution, and σ_d the standard deviation of d_i is estimated from the sample data, the ratio between the mean $\sum_i d_i / N$ and the standard error σ_d / \sqrt{N} follows a t distribution with $N - 1$ degrees of freedom under the null hypothesis. The normality assumption might be problematic, however, because d_i are formed by subtracting probabilities with limited range.

In a permutation test, the distribution of the test statistic is generated by means of simulation. The basic idea of a permutation test is that if $F_{AV}(t) = F_A(t) + F_V(t)$ holds, the realizations of the left hand side, $\hat{G}_{AV}(t)$, and on the right hand side, $\hat{G}_A(t) + \hat{G}_V(t)$ are exchangeable. In other words, under the null hypothesis, the sign of the observed race model differences $d_i = \hat{G}_{AV}(t) - \hat{G}_A(t) - \hat{G}_V(t)$ is random in each participant (this assumption is discussed in more detail in the Appendix). This exchangeability property is used to generate the distribution of the test statistic under H_0 . In each simulation, the subject-specific d_i are randomly signed, $d_i^* = s_i^* \times d_i$ (with $s_i^* = +1$ or -1 with probability $1/2$), and the simulated test statistic is calculated as $t^* = (\sum_i d_i^* / N) / (\sigma_d^* / \sqrt{N})$. This procedure is repeated, for example, 10,001 times, resulting in a distribution of 10,001 t^* values. Comparable to a standard one-sided significance test, the race model is rejected at $P \leq .05$ if the observed t_{obs} is greater than 95% of the simulated t^* .

Generalization of the permutation test for multiple time points is straightforward. I illustrate this for three time points, 220, 240, and 260 msec, resulting in three race model differences: $d_{i220} = \hat{G}_{AV}(220) - \hat{G}_A(220) - \hat{G}_V(220)$; $d_{i240} = \hat{G}_{AV}(240) - \hat{G}_A(240) - \hat{G}_V(240)$; and $d_{i260} = \hat{G}_{AV}(260) - \hat{G}_A(260) - \hat{G}_V(260)$ for each participant i , and resulting also in three partial test statistics: t_{220} , t_{240} , t_{260} . Different combination rules exist to combine the partial test statistics (Blair, Higgins, Karniski, & Kromrey, 1997). For example, the maximum can be chosen, $t_{\text{max}} = \max(t_{220}, t_{240}, t_{260})$, a special case of the “Tippett combining function” (Pesarin, 2001). For permutation, it is assumed again that, under the race model, the empirical cumulative distributions $\hat{G}_{AV}(t)$ and $\hat{G}_A(t) + \hat{G}_V(t)$ are exchangeable. A single multiplier $s_i^* = \pm 1$ is, therefore, chosen at random for each participant i and multiplied with the difference values for each time point, $d_{i220}^* = s_i^* \times d_{i220}$, and $d_{i240}^* = s_i^* \times d_{i240}$, and $d_{i260}^* = s_i^* \times d_{i260}$. Separate t^* are then calculated for the three time points, and the largest of the three t^* values is chosen, $t_{\text{max}}^* = \max(t_{220}^*, t_{240}^*, t_{260}^*)$. The race model is rejected at a one-sided $\alpha = .05$ if the observed t_{max} is greater than 95% of the simulated t_{max}^* values.

If each partial test statistic is unbiased and consistent, any combination rule that is an increasing function of each individual t value is suitable for aggregating the partial test statistics (Pesarin, 2001). In the case of testing multiple hypotheses, the t_{max} statistic follows naturally from an inference logic in which the overall null hypothesis is rejected if *any* single partial test statistic is significant. On the other extreme, the t_{min} statistic reflects an inference logic in which the global alternative is accepted only if *all* partial tests are significant. Kiesel et al. (2007) suggest an intermediate strategy: In order to control the Type I error when Inequality 2 is tested at multiple percentiles, the race model should be rejected only if more than one

(say, $c = 2$) of the multiple t tests are significant. In a multivariate permutation test, this corresponds to the $t_{\max(2)}$ statistic; in other words the alternative is accepted if the second greatest t value is significant.

For neuroimaging data in which the number of partial tests sometimes exceeds 1,000, test statistics adding up adjacent values exceeding a given threshold have been shown to be much more powerful than the t_{\max} statistic (e.g., E. Maris & Oostenveld, 2007). The reason for this is that brain activations are typically localized in adjacent voxels, a fact not recognized by the t_{\max} statistic. As violations of the race model prediction can be expected to be visible at more than just one single time point, we also evaluated the power of the t_{sum} statistic.

Simulations

To assess the Type I error probability and the power of the permutation test, simulations were performed using small, moderate, and large samples and combination rules [$t_{\max(c)}$, t_{sum}]. In the first simulation, Type I error probability was assessed using data simulated by a race model. Power to detect violations of the race model was assessed using data simulated by a coactivation model. Detection times for auditory and visual stimuli (\mathbf{D}_A , \mathbf{D}_V) were sampled from inverse Gaussian distributions (barrier $a = 100$, drift $\mu_A = 1.34$, $\sigma_A = 11.7$, $\mu_V = 0.53$, $\sigma_V = 4.3$; see, e.g., Schwarz, 1994, Table 1). Under the race model assumption, RTs for auditory–visual stimuli were generated from the pairwise minima of antithetic pairs of auditory and visual detection times. “Antithetic” means that \mathbf{D}_A and \mathbf{D}_V were not sampled independently but that small realizations of \mathbf{D}_A were paired with large realizations of \mathbf{D}_V , and vice versa. This induces a negative correlation between the component detection times, which maximizes the redundancy gain (Miller, 1986, p. 337). Observable RTs were generated by adding an independent normally distributed motor component to the detection time ($\mu_M = 160$, $\sigma_M = 21.7$; see Schwarz, 1994, Table 1). To account for individual differences in detection time and response speed, observer-specific parameters μ_A^* , σ_A^* , μ_V^* , σ_V^* , μ_M^* , σ_M^* were sampled from normal distributions around the values described [$\mu_A^* \sim N(\mu_A, 0.1 \mu_A)$, $\sigma_A^* \sim N(\sigma_A, 0.1 \sigma_A)$, etc.].

The second simulation is based on bootstrap samples of RTs from a published experiment (Gondan, Lange, Rösler, & Röder, 2004). Fourteen participants took part in this experiment; each participant made 200 speeded responses to auditory, visual, or auditory–visual stimuli. Small data sets were simulated by sampling, with replacement, $N = 10$ participants from the entire group of participants. Within each selected participant, RTs for unimodal stimuli were generated by sampling $n = 20$ auditory and visual RTs. Under the race model assumption, RTs for auditory–visual stimuli were generated again by antithetic bootstraps from the auditory and visual RTs. For the coactivation model, auditory–visual RTs were generated by resampling from the observed auditory–visual RT distribution. This implicitly assumes that coactivation actually occurred in the experiment, which seems reasonable, since classical analysis using multiple t tests indicated substantial violations of the

Table 1
Type I Error and Power of the Permutation Test and of Multiple t Tests

Type I Error and Power of the Permutation Test and of Multiple t Tests																		
Small Samples [†]						Moderate Samples ^{††}						Large Samples ^{†††}						
$c = 1$	$c = 2$	$c = 3$	$c = 4$	$c = 5$	t_{sum}	$c = 1$	$c = 2$	$c = 3$	$c = 4$	$c = 5$	t_{sum}	$c = 1$	$c = 2$	$c = 3$	$c = 4$	$c = 5$	t_{sum}	
(A) Simulated Data																		
Race Model (Type I Error)																		
Permutation t_{\max}	.042	.043	.047	.045	.054	.041	.048	.049	.048	.048	.048	.050	.048	.053	.047	.048	.046	
Multiple t tests	.126	.081	.041	.027	.017	.135	.077	.041	.029	.014		.129	.081	.045	.031	.012		
Diffusion Superposition Model (Power)																		
Permutation t_{\max}	.410	.432	.442	.455	.431	.838	.880	.869	.869	.835	.885	.997	.998	.998	.997	.992	.999	
Multiple t tests	.641	.538	.438	.331	.241	.940	.901	.843	.765	.660		.999	.999	.995	.988	.961		
(B) Bootstrap Samples From Gondan et al. (2004)																		
Race Model (Type I Error)																		
Permutation t_{\max}	.034	.038	.037	.028	.003	.048	.047	.042	.047	.052	.049	.048	.052	.052	.050	.046	.054	
Multiple t tests	.123	.077	.047	.025	.016	.122	.077	.042	.025	.013		.136	.074	.041	.029	.016		
Original Data																		
Permutation t_{\max}	.542	.469	.323	.021	.000	.955	.962	.942	.773	.209	.957	.999	.999	.999	.997	.983	1.00	
Multiple t tests	.895	.840	.802	.716	.639	.997	.994	.986	.974	.948		1.00	1.00	.999	.999	.996		

Note.—Amount of simulations in which a p value below .05 was observed— c : number of percentile-specific partial tests required to be significant. [†](A), (B): $N = 10$, $n = 20$. ^{††}(A) $N = 20$, $n = 30$; (B) $N = 14$, $n = 40$. ^{†††}(A) $N = 30$, $n = 50$; (B) $N = 14$, $n = 100$.

race model prediction (Gondan et al., 2004, Figure 2A). Samples of moderate size were generated with $N = 14$ and $n = 40$; large samples, with $N = 14$ and $n = 100$.

Multiple t tests were carried out as described by Kiesel et al. (2007): For the auditory–visual RT distribution, percentiles 5, 10, 15, 20, 25, and 30 were estimated as recommended by Hyndman and Fan (1996). The sum of the auditory and the visual RT distribution was estimated by first mixing the two distributions [this yields an estimate of $\frac{1}{2} F_A(t) + \frac{1}{2} F_V(t)$], then dropping the slower 50% of the responses [this yields an estimate of $F_A(t) + F_V(t)$]. For each percentile, a paired t test was calculated. The race model was rejected if a significant ($\alpha = 5\%$, one sided) violation of Inequality 2 was observed at $c = 1, 2, 3, 4, 5$ percentiles (not necessarily adjacent).

The permutation test was performed at percentiles 5, 10, 15, 20, 25, and 30 of the participant-specific mixture of RTs for A, V, and AV. At each percentile, partial t values were calculated using the observed proportions of RTs below t for A, V, and AV. Two combining functions were evaluated: The first function corresponds to the $t_{\max(c)}$ statistic described above, with c taking values between 1 and 5 (i.e., the greatest, second, third, fourth, and fifth t value was chosen). The second function, t_{sum} , linearly adds the partial t values exceeding 1.64 (not necessarily at adjacent time points).

Under the race model assumption, significant violations ($P \leq .05$) of the race model should be observed in about 5% of the simulations. In line with the simulation results of Kiesel et al. (2007), t tests at multiple percentiles lead to Type I error probabilities higher than the nominal α level. This inflation of the Type I error was observed for all sample sizes (e.g., 12.4% instead of 5% in large samples; see Table 1A). Requiring more than one partial test to be significant somehow protects the Type I error, but the target α level is never attained. In contrast, the permutation test always controls the α level for all sample sizes, regardless of the combination rule used. Very similar results were obtained for bootstrap samples from the Gondan et al. (2004) study (Table 1B). Here, for high values of c , however, the permutation test is overly conservative in small samples.

The power to detect coactivation was assessed using RTs for auditory–visual stimuli generated by the diffusion superposition model (Schwarz, 1994). The superposition model assumes that the diffusion parameters of the auditory and the visual channel add up linearly ($\mu_{AV} = \mu_A + \mu_V$, $\sigma_{AV}^2 = \sigma_A^2 + \sigma_V^2$) in redundant signals. Consequently, the criterion for detection is reached earlier. Table 1A shows that the power of the different variants of the permutation test is about 45% for small samples and about 85% for samples of medium size ($N = 20$ participants, $n = 30$ trials per condition). The highest power was observed in t_{sum} ; that is, if the combined test statistic corresponds to the sum of the significant partial t values. Reanalysis of bootstrap samples from Gondan et al. (2004) shows similar results (Table 1B): The power of the permutation test attains 95% already for moderate samples ($N = 14$, $n = 40$). Here, the power of the t_{sum} statistic was slightly lower than for the t_{\max} statistic, and in small samples, the

power of t_{\max} is decreased if too many partial tests are required to be significant.

Discussion

In the present article, I propose a permutation procedure for testing the race model inequality in group data. The permutation test is a straightforward extension of the commonly used t test for percentiles, which has been demonstrated to be anticonservative if the trial number is low, or if tests are performed at multiple percentiles (Kiesel et al., 2007). Permutation procedures are known to control the Type I error in correlated significance tests; they have repeatedly been used for testing differences in fMRI or EEG studies (Blair & Karniski, 1993; E. Maris & Oostenveld, 2007). Since the distribution of the test statistic under the null hypothesis is generated by means of simulation using the observed data, permutation tests do not rely on the normality assumption (see, e.g., Hesterberg, Moore, Monaghan, Clipson, & Epstein, 2005). Although the test statistics and the P values of permutation tests are conditional on the observations, it can be shown, that permutation tests control the Type I error unconditionally, if sampling is unbiased (E. Maris & Oostenveld, 2007). Percentile functions need not be estimated for the test; the permutation test of the race model inequality should therefore, not be affected by biases described by Kiesel et al. (2007). Problems related to multiple testing are controlled because the test statistic is aggregated over the time points before its sampling distribution is determined.

Particularly in group studies with small sample sizes, researchers might choose to first test the race model prediction within each participant (e.g., G. Maris & Maris, 2003; Miller, 1986; Vorberg, 2008). In a second step, the participant-specific P values (P_i) can be integrated using meta-analysis (e.g., the Fisher omnibus test, $-2 \sum_i \log P_i \sim \chi_{df=2N}^2$, with N denoting the number of participants). In such a single-participant analysis, the null hypothesis states that the RTs for each participant follow the race model. The meta-analysis then examines whether the race model holds in *every* participant; rejection of this hypothesis implies that the race model is violated in at least 1 participant.

In the t test or permutation test, the null hypothesis is formulated for the entire group, stating that the race model holds, *on average*, for the group of participants. More specifically, the pairs $\langle \hat{G}_{AV}(t), \hat{G}_A(t) + \hat{G}_V(t) \rangle$ are exchangeable over the subject population. This group hypothesis includes the subject-specific hypotheses: If the race model holds for all participants, it holds, on average, as well. Whereas the meta-analysis of the subject-specific null hypotheses will presumably be more powerful than the test the group hypothesis, rejection of the group hypothesis implies that the race model is violated by the average participant. Some researchers might consider such an alternative more interesting than the information that the race model is violated in at least one participant.

Just as the race model inequality is a rather conservative upper bound (Miller, 1982; Ulrich & Giray, 1986), significance tests for detecting violations of the race

model are conservative, as well. Small coactivation effects might not be detected. I illustrated the permutation test using three fixed values of t (220, 240, 260 msec). Of course, the permutation test is not limited to these three arbitrary time points. Instead of three fixed values for t , the race model differences d_i can be determined at prespecified percentile estimates, for example, $t_{.10} = \hat{G}^{-1}(.10)$, $t_{.15} = \hat{G}^{-1}(.15)$, and so on, with $\hat{G}(t)$ denoting the participant-specific mixture of the three distributions $\hat{G}(t) = [\hat{G}_A(t) + \hat{G}_V(t) + \hat{G}_{AV}(t)]/3$. The RTs corresponding to these percentiles can be determined separately for each participant. While this is being done, interindividual differences in response speed are taken into account, an important advantage of the paired t test for percentiles described above. For high values of t , the three distributions approach their asymptote 1.0, and $F_{AV}(t)$ is always below $F_A(t) + F_V(t)$. Therefore, only the lower percentile range should be used for testing the race model (e.g., up to percentile 25; see Kiesel et al., 2007). Theoretical considerations or pilot data can be used to narrow down the range of percentiles used for the test (Kiesel et al., 2007, p. 549). Of course, the permutation test only protects the Type I error for a given set of percentiles defined a priori. Choosing the dependent variable after data inspection generally leads to biased decisions.

In many experiments, it is possible to anticipate the onset of the next upcoming stimulus, and highly motivated participants might actually press the response button before they notice the stimulus. Eriksen (1988) noted that such fast guesses bias the test of the race model. This bias is avoided by inserting catch trials in which the RT to “no stimulus” events is recorded and eliminating, for each “response” to a catch trial, a similar RT from F_A , F_V , and F_{AV} (“kill-the-twin” correction; see also Gondan & Heckel, 2008). Similarly, Miller (2004, Appendix A) shows that excluding misses from the analysis biases the test of the race model inequality again in favor of the race model.

The permutation test described in the present article may serve as a valuable tool for research on cognitive architecture, enabling powerful tests of the race model prediction at the desired Type I error level. Short programs written in R statistical language (R Development Core Team, 2009) and MATLAB (R2007b, The MathWorks, Munich, Germany) are available as online supplemental materials.

AUTHOR NOTE

This research was supported by the German Research Foundation (DFG, GO 1855/1-1). The author thanks Andrea Kiesel, Steven Blurton, and Mark W. Greenlee for help during preparation of the manuscript. Correspondence concerning this article should be addressed to M. Gondan, Department of Psychology, University of Regensburg, D-93050 Regensburg, Germany (e-mail: matthias.gondan@psychologie.uni-regensburg.de).

Note—Accepted by the previous editorial team, when John H. Krantz was Editor.

REFERENCES

- BLAIR, R. C., HIGGINS, J. J., KARNISKI, W., & KROMREY, J. D. (1997). A study of multivariate permutation tests which may replace Hotelling's T^2 test in prescribed circumstances. *Multivariate Behavioral Research*, **29**, 141-163. doi:10.1207/s15327906mbr2902_2
- BLAIR, R. C., & KARNISKI, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, **30**, 518-524. doi:10.1111/j.1469-8986.1993.tb02075.x
- ERIKSEN, C. W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception & Psychophysics*, **44**, 191-193.
- GILCHRIST, W. G. (2000). *Statistical modelling with quantile functions*. Boca Raton, FL: Chapman & Hall/CRC.
- GONDAN, M., & HECKEL, A. (2008). Testing the race inequality: A simple correction procedure for fast guesses. *Journal of Mathematical Psychology*, **52**, 322-325. doi:10.1016/j.jmp.2008.08.002
- GONDAN, M., LANGE, K., RÖSLER, F., & RÖDER, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review*, **11**, 307-313.
- HESTERBERG, T. C., MOORE, D. S., MONAGHAN, S., CLIPSON, A., & EPSTEIN, R. (2005). *Bootstrap methods and permutation tests*. New York: Freeman.
- HYNDMAN, R. J., & FAN, Y. (1996). Sample quantiles in statistical packages. *American Statistician*, **50**, 361-365. doi:10.2307/2684934
- KIESEL, A., MILLER, J., & ULRICH, R. (2007). Systematic biases and Type I error accumulation in tests of the race model inequality. *Behavior Research Methods*, **39**, 539-551.
- MARIS, E., & OOSTENVELD, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, **164**, 177-190. doi:10.1016/j.jneumeth.2007.03.024
- MARIS, G., & MARIS, E. (2003). Testing the race model inequality: A nonparametric approach. *Journal of Mathematical Psychology*, **47**, 507-514. doi:10.1016/S0022-2496(03)00062-2
- MILLER, J. [O.] (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, **14**, 247-279. doi:10.1016/0010-0285(82)90010-X
- MILLER, J. [O.] (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, **40**, 331-343.
- MILLER, J. [O.] (2004). Exaggerated redundancy gain in the split brain: A hemispheric coactivation account. *Cognitive Psychology*, **49**, 118-154. doi:10.1016/j.cogpsych.2003.12.003
- PESARIN, F. (2001). *Multivariate permutation tests: With applications in biostatistics*. New York: Wiley.
- RAAB, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, **24**, 574-590.
- R DEVELOPMENT CORE TEAM (2009). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- SCHWARZ, W. (1994). Diffusion, superposition, and the redundant-targets effect. *Journal of Mathematical Psychology*, **38**, 504-520. doi:10.1006/jmps.1994.1036
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
- TOWNSEND, J. T., & NOZAWA, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, **39**, 321-359. doi:10.1006/jmps.1995.1033
- ULRICH, R., & GIRAY, M. (1986). Separate-activation models with variable base times: Testability and checking of cross-channel dependency. *Perception & Psychophysics*, **39**, 248-254.
- ULRICH, R., MILLER, J., & SCHRÖTER, H. (2007). Testing the race model inequality: An algorithm and computer programs. *Behavior Research Methods*, **39**, 291-302.
- VORBERG, D. (2008, September). *Exact statistical tests of the race model and related distribution inequalities*. Paper presentation at the meeting of the European Mathematical Psychology Group, Graz, Austria.

SUPPLEMENTAL MATERIALS

A script in R statistical programming language (R Development Core Team, 2009) and a MATLAB script (R2007b, The MathWorks, Munich, Germany), as well as usage instructions and example data from a redundant target experiment (Gondan et al., 2004), may be downloaded from <http://brm.psychonomic-journals.org/content/supplemental>.

APPENDIX

The permutation test proposed to test the race model inequality is derived from a paired samples testing problem (see, e.g., Pesarin, 2001, ch. 2). In a typical paired sample problem, each participant provides two values: \mathbf{X}_A and \mathbf{X}_B . The null hypothesis states that the two values have the same distribution, $H_0: \mathbf{X}_A \sim \mathbf{X}_B$ (with \sim denoting “equal in distribution”). This hypothesis is tested using the difference values $\mathbf{D} = \mathbf{X}_A - \mathbf{X}_B$ for each participant. If H_0 holds, these differences are symmetrically distributed around zero (Pesarin, p. 20):

$$\begin{aligned} P\{\mathbf{X}_A - \mathbf{X}_B < k\} &= P\{\mathbf{X}_B - \mathbf{X}_A < k\} \\ P\{\mathbf{D} < k\} &= P\{-\mathbf{D} < k\} \end{aligned}$$

This exchangeability property allows for generating the permutation distribution of the test statistic under H_0 by randomly signing the realizations of \mathbf{D} observed in the sample.

For the race model, we first note that the left hand side of (1) cannot exceed unity, so that the race model inequality is restated as $F_{AV}(t) \leq F_{A+V}(t)$, with $F_{A+V}(t) = \min[1, F_A(t) + F_V(t)]$ (e.g., G. Maris & Maris, 2003, Ineq. 6). For the statistical test, the null hypothesis states that $F_{AV}(t) = \min[F_A(t) + F_V(t)]$, with $F_c(t) = P\{\mathbf{T}_c \leq t\}$ denoting the probability for an RT below a given t in condition $c = A, V, AV$. This implies that over the subject population, the pairs of empirical cumulative RT distributions $\hat{G}_{AV}(t)$ and $\hat{G}_{A+V}(t) = \min[1, \hat{G}_A(t) + \hat{G}_V(t)]$ are exchangeable. Without loss of generality, assume that the number of trials n is equal in the three conditions. Therefore, H_0 can be simplified to $N_{AV}(t) = N_{A+V}(t) = \min[n, N_A(t) + N_V(t)]$, with $N_c(t) = F_c(t) \times n$ denoting the expected number of RTs below t . If $F_c(t)$ is constant during the experiment, the number \mathbf{X}_c of responses actually observed to be smaller than t follows a binomial distribution (see Townsend & Nozawa, 1995, Appendix C):

$$\begin{aligned} \mathbf{X}_{AV} &\sim \text{Bi}[n, F_{AV}(t)] \\ \mathbf{X}_A &\sim \text{Bi}[n, F_A(t)] \\ \mathbf{X}_V &\sim \text{Bi}[n, F_V(t)] \end{aligned}$$

In general, $\mathbf{X}_A + \mathbf{X}_V$ does not follow a binomial distribution, and neither does $\mathbf{X}_{A+V} = \min[n, \mathbf{X}_A + \mathbf{X}_V]$. Consequently, \mathbf{X}_{AV} does not have the same distribution as \mathbf{X}_{A+V} , calling into question whether $\mathbf{D} = \mathbf{X}_{AV} - \mathbf{X}_{A+V}$ is symmetrically distributed around zero. Therefore, strictly speaking, it is not possible to derive a permutation argument for the test of the race model inequality.

However, if the number of trials n is sufficiently large, the three binomial distributions converge to normal distributions because of the central limit theorem, $\mathbf{X}_c \sim N[n F_c(t), n F_c(t) S_c(t)]$, with $S_c(t) = 1 - F_c(t)$ denoting the survivor function for condition c . A similar argument holds for the observed proportions of RTs below t ,

$$\begin{aligned} \hat{G}_{AV}(t) &\sim N[F_{AV}(t), F_{AV}(t) S_{AV}(t)/n], \\ \hat{G}_A(t) &\sim N[F_A(t), F_A(t) S_A(t)/n], \\ \hat{G}_V(t) &\sim N[F_V(t), F_V(t) S_V(t)/n]. \end{aligned}$$

Sums and differences of normal distributions are again normally distributed such that, under the null hypothesis, the race model difference for the i th participant, $d_i = \hat{G}_{AV}(t) - \hat{G}_{A+V}(t) = \hat{G}_{AV}(t) - \min[1, \hat{G}_A(t) + \hat{G}_V(t)]$ can be assumed to be more or less normally (and, thus, symmetrically) distributed around zero. Therefore, exchangeability of d_i is established if the number of trials n is large enough.

It should be pointed out that the number of trials n —not the number of participants—is crucial for the normal approximation. Since the test statistic t_{obs} involves the sum of d_i over participants, biases due to insufficient normal approximation might even accumulate. The simulations (Table 1), however, suggest that the permutation test is Type I-safe, even for experiments with only 20 trials per condition.

As in most race models $F_{AV}(t) < F_A(t) + F_V(t)$ —for instance, the independent channels model (Miller, 1982, p. 253)—the observed d_i will tend to negative values. In this case, the permutation procedure overestimates the amount of positive d_i^* when generating the permutation distribution of the test statistic t^* . To reject H_0 at a nominal significance level of $\alpha = .05$, the observed t_{obs} must be greater than 95% of the t^* . Therefore, as for most race model tests, the permutation test is conservative, not fully exploiting the prespecified α level.